Core Network Design for IPTV Transport

John Evans, Distinguished Consulting Engineer
john@cisco.com
Core Network Design for IPTV Transport

- Reference Architecture
- IP Transport Options
- Video SLA Requirements
- Unicast Routing Protocol Design
- Multicast Architecture
- Resiliency
- QOS Architecture
- VQE
- Service Management and Monitoring
IPTV Core Network Reference Architecture

Super Head End (SHE)

DCM / VQE

MSE

Source

Encoder

Source

Encoder

DCM

Super Head End (SHE)

PoP

Access and Aggregation Network

Video Hub Office (VHO) / Regional Head-end (RHE)

Encoder

Source

DCM / VQE

MSE

Video Service Office (VSO)

Home Network

Access and Aggregation Network

Core Error Domain

SHE Error Domain

Access and Aggregation Error Domain

Core IP / MPLS Network
**SUPER HEAD END**

- SHE is where common or “national” channels are acquired.
- There are typically two SHEs for resiliency.
- The functions within the SHE include:
  - The encoders that encode the received video signal to MPEG, e.g., Cisco D9032 (MPEG-2 SD), D9050 (MPEG-2 HD), D9034 (H.264 SD), D9054 (H.264 HD).
  - Video multiplexers, which are responsible for muxing / demuxing the encoded video streams into the format required for the common channel line-up, e.g., D9900 Digital Content Manager (DCM).
REGIONAL HEAD END

- The RHE is where local channels are acquired and then added to the common channel line-up to create the channel line-up for a given region.
- The components within the RHE are typically a subset of those within the SHE.
- If the national channel line-up needs to be customized for each region, video-multiplexing equipment at the RHE (e.g. D9900 DCM).
**IPTV Core Network Reference Architecture**

**VIDEO HUB OFFICE (VHO)**
- Where most of the distributed video application components reside.
- Many video application components have a centralized component (e.g., CDS Vault) and a distributed component (e.g., CDS streamer).
- The VHO is typically located at the boundary between the core and access/aggregation networks.
IPTV Core Network Reference Architecture

POINT OF PRESENCE (POP)
• Where the termination devices for IP services are located.

VIDEO SERVICE OFFICE (VSO)
• Where the aggregation nodes (and sometimes access nodes) are typically located.
• The VSO is typically the closest provider-owned facility to the subscriber, e.g., the central office or local exchange.
• This may be where the Cisco Visual Quality of Experience (VQE) capability may be located.

Access and Aggregation Error Domain

Core Error Domain
Core Video Transport: Primary Technology challenges

- The Primary Technology and Competitive Market Challenges:

1. Basic transport
   How to shift the packets … IP or MPLS, native or VPN?

2. Video service SLA
   How to ensure that the IP network delivers the required SLAs

3. Service Monitoring and Management
   How to verify that the IP network is delivering the required SLAs for video, and to identify problem areas
p2mp Transport Options

- IP
  1) Native
  2) mVPN, i.e. mGRE

- LSM
  3) p2mp TE global *
  4) MLDP global **
  5) MLDP ** + mVPN

* LSP scalability issues make TE a non starter for bidirectional (mp2mp) mVPN

** Can be used a) without TE FRR or b) in conjunction with TE FRR
# Service Requirements Summary

<table>
<thead>
<tr>
<th></th>
<th>Video Contribution</th>
<th>Video Secondary Distribution</th>
<th>Managed Enterprise mVPN</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PIM mode</strong></td>
<td>SSM only</td>
<td>SSM only</td>
<td>SM and SSM</td>
</tr>
<tr>
<td><strong>#S per G</strong></td>
<td>1 or 2</td>
<td>1 or 2</td>
<td>1 or 2</td>
</tr>
<tr>
<td><strong>#(S, G)</strong></td>
<td>#(S, G) &lt; 1000</td>
<td>#(S, G) &lt; 1000</td>
<td>100s (S, G) per VPN; 100s of VPNs</td>
</tr>
<tr>
<td><strong>Receivers per G</strong></td>
<td>&lt;10</td>
<td>Millions</td>
<td>100s of sites; potentially 1000s</td>
</tr>
<tr>
<td><strong>MDT dynamism</strong></td>
<td>100s of new trees per day; trees static once established</td>
<td>Static trees</td>
<td>MDT is dynamic; joins and leaves may impact core</td>
</tr>
<tr>
<td><strong>FRR or equivalent</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Admission control and bw reservation</strong></td>
<td>Yes (time limited reservations)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td><strong>Offload routing</strong></td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td><strong>Path diversity</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>mVPN requirement</strong></td>
<td>No</td>
<td>Possibly – for wholesale services</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>p2mp or mp2mp?</strong></td>
<td>p2mp</td>
<td>p2mp</td>
<td>mp2mp</td>
</tr>
</tbody>
</table>
### IP / MPLS p2mp technology choices

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Plain IP Multicast</th>
<th>p2mp MPLS TE</th>
<th>mLDP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Convergence</strong></td>
<td>&lt; ~1s</td>
<td>~50ms</td>
<td>&lt; ~1s</td>
</tr>
<tr>
<td></td>
<td>(link and node failures)</td>
<td>(link failures only)</td>
<td>(~50ms with p2p MPLS TE FRR LP)</td>
</tr>
<tr>
<td><strong>Offload routing</strong></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>(IGP metric based traffic engineering)</td>
<td>(IGP metric based traffic engineering)</td>
<td></td>
</tr>
<tr>
<td><strong>Path separation</strong></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>(MoFRR or MTR)</td>
<td></td>
<td>(MoFRR or MTR)</td>
</tr>
<tr>
<td><strong>Admission control and bw reservation</strong></td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td></td>
<td>(RSVP)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Scalable mp2mp MVPN</strong></td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td><strong>Typical Application</strong></td>
<td>Secondary Distribution</td>
<td>Contribution</td>
<td>Enterprise VPN</td>
</tr>
</tbody>
</table>
Video SLA Requirements

### SLA Requirements

- **Availability targets of 99.999%**
- **One-way end-to-end delays $\ll 100\text{ms}$**
- **Minimise packet loss**

### Network Design Solution

- Resilient network topology, Resilient sources, Carrier class platforms
- Deploy Diffserv IP QOS
- Fast Routing Protocol Convergence, Anycast source, Cisco VQE
Unicast Routing Design

- The Interior Gateway routing Protocol (IGP) can be OSPF or ISIS
- It is recommended that a single IGP area/level be used within the core
- It is recommended that routes to the multicast sources be carried in the IGP
  - There is no benefit in carrying routes to the source in BGP
- It is recommended that fast routing protocol convergence is used
Multicast Routing Design

- PIM-SSM is recommended as the multicast routing protocol

- **Source Specific Multicast (SSM)**
  SSM [RFC4607] is the IP multicast model used where the receiver (or some proxy) sends (S,G) joins rather than (*,G) joins, to indicate that it wants to receive traffic sent by a specific source S to a group G

- **Any Source Multicast (ASM)**
  By contrast to SSM, in the Any Source Multicast (ASM) model, a receiver expresses interest in all traffic destined to a multicast group G, hence the multicast network must discover all multicast sources sending to that address and forward data from all sources to all interested receivers
Multicast Routing Design

Source Specific Multicast (SSM)

- With SSM, the network builds a separate distribution tree for each multicast source and clients immediately receive content directly from the source.

- PIM-SSM is the subset of PIM-SM [RFC 4601] used to create (S,G) trees.

- There is no shared tree (*,G) state, nor Rendezvous Points in SSM.

- SSM requires that multicast receivers explicitly specify a source address in multicast group join requests, e.g., which can be supported with IGMPv3 or through mapping IGMPv2 join requests to specific sources.
Multicast Routing Design

Why Source Specific Multicast (SSM)?

- SSM offers benefits over ASM for one-to-many dissemination-style applications, such as IPTV with one or more senders whose identities are known before the application sessions start:
  1. **Simplification.** SSM relieves the network of needing to discover where multicast sources are present and removes the need for RPs
  2. **Security.** SSM is considered more secure than ASM, because the multicast clients must know both the multicast group and source addresses in order to join the group. Further, other devices acting as sources transmitting to the same multicast group are not able to interfere with the service, which could happen with ASM.
  3. **Scalability.** SSM offers better scalability
     a) SSM can reduce the amount of multicast state information that the network must maintain
     b) SSM improves scaling for multicast group addressing as the total number of group addresses are effectively multiplied by the total number of usable source IP addresses in the provider’s network.
IP Video Service Optimization: Multicast Traffic Replication

- Provide TV quality video services over a converged IP core network
- Innovative Service Intelligent Fabrics built for massive video traffic (multicast) replication and forwarding
- Superior In-fabric replication (compared to inefficient ingress card replication in competitors)
Resiliency

Core network link or node failures.

- The core network is built resiliently with no single points of failure, so that on the failure of any single component (link or node) between the head-end and the receivers, an alternate path will be available.

- On the failure of a core link or node, there will be a loss in connectivity until the IGP reconverges on the alternate path.

- Unicast and multicast fast convergence are used to minimise the loss of connectivity
Fast IP Routing Protocol Convergence

- Fast IP routing protocol convergence reduces the loss of connectivity following network (link and node) failure cases.

- Multicast convergence is dependent upon unicast convergence due to the requirements of the Reverse Path Forwarding (RPF) check by the multicast control plane to build the source tree.

- Therefore, fast unicast convergence is a prerequisite for fast multicast convergence.

- Implementation and protocol optimisations that can deliver sub second convergence times for unicast (OSPF, ISIS, BGP) and multicast.
Multicast (SSM) Fast Convergence (CRS-1)
Convergence Time for All Channels Following a Network Failure

No more than one I-frame loss even in the worst case

Prefix prioritisation allows important groups (e.g. premium channels) to converge first

Tested with 2500 IGP prefixes and 250k BGP routes
Impact of multicast convergence on unicast convergence is negligible
Resiliency

Source failure:

- Multiple active sources are used for resiliency
- On failure of the active source, anycast (or “priority cast”) source is used in order to switch to the alternate source
- Anycast relies upon each source using the same (S, G) for the same channel; each receiver joins to the topologically closest source for that group.
  
  On the failure of one source or the link connecting to it, the subnet will no longer be available and the network will reconverge on the working source.

- With “priority cast”, each receiver joins with the highest priority DCM for that group, where the priority is influenced either by using different prefix lengths or by engineering IGP metrics.
Anycast Source

(S1, G1)

I will send join to the nearest S1

IGMP Report

STB

(S1, G1)

I will send join to the nearest S1

IGMP Report

STB
Anycast Source

I will send join to the nearest S1

(S1, G1)

(S1, G1)

IGMP Report

STB

IGMP Report

STB
QOS Design

- Diffserv IP QoS mechanisms are used to control queuing delays and ensure that video control messages and MPEG packets are not dropped.
- Diffserv allows us to support multiple classes of traffic with different under- or over-provisioning ratios, hence SLAs, per class of service.
- Provides isolation between services – mitigating the risk associated with failures when supporting differential services.
- With an appropriate QOS design and capacity planning, the only network events that should result in a visual impairment to the video service are packet losses due to lower-layer errors or network element failures.

- Key design considerations
  - How many classes are needed?
  - What marking scheme to use?
  - Scheduler configuration.
Core Classes – How many is enough?

- At the network edge different classes are used for both SLA differentiation and for application isolation.
- In the core, typically where traffic is aggregated, supporting the QoS requirements for a class can effectively be translated into a problem of bandwidth provisioning.
  - Different classes are used for SLA differentiation and service isolation
    e.g. isolation from DDOS attacks
- Hence, not necessarily same number of classes in core as at the edge
  - May be a many-to-one mapping from edge-to-core
Core Classes of Service

Example 5 class core model

- **VoIP.** This class supports residential and business VoIP services.

- **Video broadcast (Vid).** This class targets real-time video streaming applications such as IPTV and Cable TV, which have requirements for low delay, low jitter, and low loss.

- **Business Data Class (Bus-in / Bus-out).** This class supports Business VPN services, which differentiate between in- and out-of-contract traffic, where a loss commitment is provided for in-contract traffic, and out-of-contract traffic is able to re-use unused bandwidth from other classes within the same policy.

- **Network Control Class (Ctrl).** The Ctrl class is dedicated for network control traffic, ensuring that bandwidth on the core links is guaranteed for essential functions typically including routing protocols and for network management traffic, such as Telnet or SNMP access.

- **Standard Class (Std).** The Std class is used for all other data traffic, i.e., all other traffic that has not been classified as VoIP, Vid, Bus or Ctrl. This may be used for Residential Internet services or for Business VPN services with a lower SLA than the Bus class.
## Classification and Marking scheme

<table>
<thead>
<tr>
<th>Service</th>
<th>PHB</th>
<th>DSCP</th>
<th>EXP</th>
</tr>
</thead>
<tbody>
<tr>
<td>VoIP / Vid. conf</td>
<td>EF</td>
<td>46, CS5 (40)</td>
<td>5</td>
</tr>
<tr>
<td>Network control</td>
<td>AF</td>
<td>CS6 (48)</td>
<td>6</td>
</tr>
<tr>
<td>Video (Broadcast)</td>
<td>AF41</td>
<td>34, CS4 (32)</td>
<td>4</td>
</tr>
<tr>
<td>VoD (Real-time)</td>
<td>AF42, AF43</td>
<td>36, 38</td>
<td>3</td>
</tr>
<tr>
<td>Bus (in-contract)</td>
<td>AF</td>
<td>CS2 (16)</td>
<td>2</td>
</tr>
<tr>
<td>Bus (out-of-contract)</td>
<td>AF</td>
<td>CS1 (8)</td>
<td>1</td>
</tr>
<tr>
<td>Standard</td>
<td>Default</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Video optimised Diffserv Schedulers

- Cisco leads the industry in the development and support of multi-priority schedulers implementations
- Enables differentiation between premium services, requiring bounded delays
Industry leading IP Diffserv Implementation

- With Cisco’s optimised IP Diffserv implementations, worst-case per hop delays $<<1\text{ms}$ for high-speed links
- Jitter of $<1\text{ms}$ is realiseable across core networks today with Cisco’s video optimised products
- See [FILSFILS] and [EVANS]

![Worst case priority queue delay](chart.png)

- Delay (µs)
  - OC3
  - OC12
  - OC48

Where OC3, OC12, and OC48 are different interface speeds.
Cisco Visual Quality Experience (VQE)

- Cisco VQE includes client and server functionality and provides the following capabilities
  - Error Repair – both retransmission-based and FEC-based
  - Quality of Experience monitoring – using RTCP receiver reports
  - Fast channel change (future)

- Both a client and server component
  - VQE-Server (VQE-S) is an appliance
    - Joins multicast channels
    - Responds to requests for VQE services (repairs, channel change) from VQE-Clients (STBs)
    - Generate per-channel real-time reception quality reports (loss, delay, jitter,...)
  - VQE-Clients (VQE-C) are part of STB software
    - Joins different channels
    - Facilitates error repair
    - Generates per-channel per-client real-time reception quality reports (loss, delay, jitter,...)

- Requires RTP-encapsulated video
  - Uses RTCP for client ➔ server communication
IPTV Error Repair – Automatic Repeat reQuest (ARQ)

- Two elements:
  - VQE Servers: deployed at the edge of the network between the receivers and access network – cache content and retransmit packets to individual receivers on demand
  - VQE STB Client: receives and buffers RTP packets, looks for dropped packets, triggers retransmission requests.
- Correction applied on an as needed basis – minimal overhead
- Provides statistics per each receiver and allows planning corrective actions
FEC adds redundancy to the transmitted data to allow the receiver to detect and correct errors (within some bound) without the need to resend any data.

- AL-FEC defined by Pro-MPEG CoP3 (SMPTE 2022)
IPTV Error Repair – Combined Solution

- Layered Repair Solution using AL-FEC and RTP Retransmission together!
- 2 complimentary technologies mitigate packet loss impact
  AL-FEC initial defense; retransmission supplementary defense
- Optimize FEC protection and overhead for majority of nodes
- Increase reach through augmentation with retransmission
- Cisco is the ONLY vendor to offer a hybrid solution
Service Management and Monitoring Video Transport Monitoring

- Transport service provider must ensure proper service from source boundary to receiver or access service boundaries

- Many types of video transport measurement metrics:
  - Loss – primary importance
  - Latency and Jitter – secondary importance

- Need techniques to isolate where and what video packets are being dropped in the network

- Multi-pronged approach
  - Active and passive per flow video transport monitoring
  - Video quality monitoring
  - Overarching video service management solution
Video Transport Measurement Metrics

- **Media Rate Variation (MRV)**
  Measurement of a flow as a variation from the expected arrival rate beyond nominal thresholds
  Discriminates between problems at the source boundary, at the edge boundary, within the network

- **Media Stop Event (MSE)**
  Boolean value representing that no packets have been received on a valid flow (cache entry exists) for measurement period. Assumes knowledge of the transport rules for valid flow delivery

- **Media Loss Rate (MLR)**
  Used by probe vendors to specifically measure MPEG packet loss (e.g. discontinuity)
  e.g. may reported by MDI loss factor (LF)
  May be required for VBR flows
CSI1  

MRV: MRV is complementary to MDI. This may only be used for fault isolation, not to indicate a fault or warning. Dave O. is doing some research to predict loss with 90% accuracy (based on the premise that after a loss there is an increase - but we need to ensure that we have the rate right).

RTP:

MLR: Discontinuity measurement. Discontinuity tied to number of PIDs...
Sample Use Case: $MDI + MRV$

*Using Expected Rate*

![Diagram showing a network with various components and connections, highlighting MDI and MRV with expected rates and a loss indicator.](image)
Cisco Video Assurance Management Solution (VAMS)

Presentation

Cisco VAMS

Cisco Info Center for Video

Rules Engine

Traps & Cross-launch

Cisco Multicast Manager

Cisco ANA

Gateway

VNE Servers

Traps

Polling/Traps

Polling/Traps Commands

Data Sources

Video Probes

Headend

Core

CRS-1 Router

CRS-1 Router

7600

4948 or 7600

Distribution

CRS-1 Router

CRS-1 Router

4948 or 7600

Aggregation

Last mile

Home

Collection

Data Sources

Video Probes

Router 7600

© 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential Presentation_ID
Additional Resources

- Design and Implementation Guide is available under NDA
  Request from account team
References


